Interaction events in contactless gestural systems: from motion to interaction

Nicolas Bremard
Universite Lille 1
Cite scientifique
59655 Villeneuve d'ascq, FRANCE
nicolas.bremard@lifl.fr

Laurent Grisoni
Universite Lille 1
Cite scientifique
59655 Villeneuve d'ascq, FRANCE
laurent.grisoni@lifl.fr

Bruno De Araujo
INRIA Lille
40, avenue Halley
59650 Villeneuve d'ascq, FRANCE
bruno.araujo@inria.fr

ABSTRACT

Contactless interactive systems appear to have an interesting context of application in interactive art, however practical example often involve low interactivity; we think this is due to the fact that this interaction paradigm is still poorly understood, and tools for interaction design are still missing. We present a simple interaction technique (*Point-and-Move*) that allows to point-and-click without instrument through distance; through this example, we show that the presented concept and framework has nice and flexible properties. Finally, we illustrate the approach on the two artistic installation we participated to, and also provide uncontrolled user experience feedback on the presented interaction technique, from these installations.

Categories and Subject Descriptors

H.5.2 [**Information Interfaces and Presentation**]: User interfaces—*Interaction styles, Prototyping*

Keywords

parameterized gesture, contactless interaction, numerical art

1. INTRODUCTION

An important challenge about contactless gestural interaction is to understand how to take the best from user gesture expressivity: WIMP paradigm is based on as-elementary-as-possible interaction input (control of a pointer position on a screen, and event generation through clicking). By comparison, user hand movements has a much larger expressivity, from which interactive applications only take poor benefit. Such richness can be, on some specific application contexts, a difficulty (e.g. within the multi-touch context, it is known that users do not matter about how many fingers actually touch the surface [17]; understanding such variability is a key point for multi-touch systems [11]). In the context of gestural interaction, the problem is still very large, since cases of use are not clearly identified yet, and many open question remains unsolved, the most important ones probably being *why interacting with a computer using hands*, and *how to achieve that.* Indeed, hand movements are classically the mean to act on physical world, complement inter-human communication, mimic something . It is a recent idea to think about user hand movements as a mean to interact with an application. The idea is all the more complex as structured tasks are foreseen.

One of the first question that is to be studied is how to handle the non-separable nature of user movements: grasping a glass, or pointing at something, involves one single gesture, and the intrinsic nature of such movement is very much different from that of grasping a mouse and moving it for interaction:from a computer science point of view, interaction gesture would often have to be interpreted in terms of action (*what is being done?*) and parameters (*on which part of the application? In which manner?*). In this article we address the question of how to generate interaction events from user gestures. We aim at providing elements that are independent of any standard recognition technique; instead, our goal is to provide a framework for fast prototyping of explicit gestural interaction recognition and design.

Few tools are currently existing for fast prototyping of contactless and full-body interactions. The context of interactive art is an interesting application context to such systems, as several recent movement low-cost acquisition systems open the way to accessible installation (however these low-cost system are often of low space and time precision). We introduce an approach based on simple, basic numerical tools, that can be easily implemented and combined by interaction designer, so that any target system can be prototyped. Our approach is flexible, can be easily adapted to specific input, and does not involve any sophisticated recognition method; as a result it can be easily reproduced and adapted to specific interaction design problems. We illustrate our method within two artistic installations where we propose an interaction technique combining pointing and selection within gestural, distant and non-instrumented, interaction context using low-cost non-intrusive tracking devices such as the Kinect depth sensor or RGB cameras. Finally, we discuss user feedback from two artistic installations.

MOCO'14, June 16-17 2014, Paris, France.
ACM 978-1-4503-2814-2/14/06 ...$15.00.
http://dx.doi.org/10.1145/2617995.2618027.

2. RELATED WORK

Researches have already been conducted to improve standard gestures recognition processes within contactless interactions. Wilson et al. ([15],[16]) proposed to improve Hiden

Markov Model (HMM) by extracting meaningful parameters from the move and modifying HMM's output with it. He demonstrated that meaningful parameters could be computed from moves and be used outside a gesture, now called a parameterized gesture.

Karam and Schraefel [8] proposed a classification of gestures based on four key elements: gesture styles, application domains, input and output technologies. They determined six categories : deictic, manipulative, semaphoric, gesticulation, sign language and multiple. It have been reviewed by Aignier et al. [1] making the distinction between static and dynamic gestures, admitting that dynamic is a key for gestures classification. While these classifications are good to describe gesture in an abstract way, they do not really describe how gestures can be decomposed into parameters to be recognized and applied on a given interaction scenario.

Several works have focused on how to define gesture within the user interaction space and discuss the idea to interact within a space that user would imagine in front of him/her. Rateau et al. [10] create this surface from a gesture, whilst Chan et al. [3] and Gustafson et al. [6] hang it to the non-dominant hand. Even if users can create as many surfaces as they need, these interactions are still in two dimension and are restraining the application with a cursor.

A special focus have been done regarding pointing gestures to a surface for emulating 2D cursor. Pierce et al.[9] use it to manipulate virtual objects. Other studies try to recognize a command, such as a click : with non dominant hand [13] or with fingers [14]. These works create a two dimensional interaction too, and are not enough for creating a high level of interactivity. Ren et al.[12] propose a three dimensional selection technique in which the hand's direction is used to emulate a command (choose an option). This technique constantly check if the hand is moving and trigger the associated command, which avoid any other gesture to be associated with this hand. Choumane et al. [4] suggest to use the dynamic, represented by acceleration, of the gesture to select an object. The technique based on acceleration let the hand free to do other gestures. That proves the use of gesture dynamic has more advantages than direction alone for gestures recognition.

Regarding tools for easing interaction design, some graphical tools have been proposed such as ICON [5] and Flowstate [2]. These tools are efficient when used with keyboards and devices with two dimensions such as mice and touch screen; but they are unusable in three dimensions : they have not been designed for it. Our proposal presents a solution to this problem easing the design of gestural based interactions using 3D data gathered from the user.

3. GESTURAL INTERACTION DESIGN FRAMEWORK

We present here a new tool for interaction design, fully compatible with contactless interactions and parameterized gestures. A parameterized gesture is a gesture from which meaningful values related to the gesture can be extracted; such as a pointed position for a pointing gesture or a direction for a manipulative action. Some of these attributes can be extracted from the gesture itself and gathered information from the tracking device without requiring any information. For example, it can be a given position, a speed, an acceleration, a direction or a duration associated to the gesture. Others attributes are related to the environment and the application context which the gesture is subject to. For example, which object (virtual or real) or person someone is pointing at while doing a pointing gesture.

3.1 General overview

Our tool is based on three components : an input device abstraction, a user model and set of observers for gesture recognition. We follow a plugin architecture to represent the input devices and forward the input data to the other components using an update method. First we update our user model converting the data gathered from the input device for example the skeleton data of tracked user when using the Kinect device. Our user model represents a full-body user by a set of joints. Each joint has three dimensional position and orientation. The set of joints can be customized , i.e. they can be ignored if the data is missing or not needed by the designer, or added if they are available by a given input device. This information is then forwarded to the observers designed to recognize a set of gestures attributes.

3.2 Observers

The Observers are basic methods, called at each input update, that evaluate an observer value. These values are tested within each interaction automata associated to interaction gestures. The following set of observers are modeled by our approach and can be combined depending on the kind of gestures they are designed to recognize.

zero-Order Observer: it detects when a given joint is close to a given state. For example it can represent the positional or angular distance to a recorded position or orientation using use absolute and relative coordinates.

First-Order Observer: these observers recognize if a joint is moving in a specific direction, according to the coordinate system, or if the speed or the angular speed of its orientation are above,below or between specified values.

Second-Order Observer: it recognizes if a joint is moving at, above, below or between specified accelerations.

Time-Switch Observer: these observers are designed to detect predefined changes about a joint. Changes can be in position, speed, acceleration, or orientations for example.

World-Aware Observers: it integrates additional informations related with the environment or the type of interaction. In most examples, they represent either physical objects or screen-displayed widgets related to a given gesture.

Personalized Observer: it allows to create your own gesture recognizer integrating it seamlessly in our architecture.

Composite Observer: it enables to combine several observers and create a small automata with them.

3.3 Gesture example: Point and Select

Providing generic interaction techniques, such as point-and-click, in contactless interaction, is a question to which no current interaction paradigms answer. Indeed, in public context, instrumented interaction is rarely available, as a result no physical button can achieve the *click* feature. Moreover, most current contactless acquisition devices are still with low resolution and frequency, making data noisy, and body reconstruction usually coarse.

In the practical artistic request described in this article, we had the two different options to explore, resulting in two way to select with different accuracies due to contrast in the gesture dynamic. These pointing techniques, proposed

Figure 1: Damassama installation setup using 24 singing bowls actuated mechanically. A Kinect camera is placed in the center underneath the first row.

Figure 3: Installation setup used by the Tempo Scaduto experience.

by [7, 4], have been implement using a Time-Switch Observer based on the acceleration and a world-aware observer combining the gesture performed by the user and the object pointed by the user in our installation:

The virtual hammer with click-and-point: In this technique, the dynamic of the gesture is done first, then pointing is used to determine where the click has been performed. This technique is efficient with big object users can point easily because they must stop their gesture on it.

The virtual gun using point-and-click: Here, users can point on entities to be selected and trigger a click at the pointing location. This technique is more accurate due to the dynamic done after pointing the position users want.

4. INTERACTIVE ART USE CASES

Our approach was demonstrated in two artistic installations created in collaboration with artists from Le Fresnoy , a digital art school from Tourcoing, France. . Both installations consist in offering to a broad audience an interactive experience where participants could interact through gestures with an artwork. Such scenario allowed to test our gestural approach in a public situation and enabled artists to take advantage of our contact-less interaction techniques. In addition, we aimed on assessing if our gestural interaction design could fit artist needs in a real artwork installation without constraining their creativity.

4.1 Damassama

The Damassama is an interactive artwork created by Leonore Mercier where the artist or the visitor can play music using a set of percussion instruments. Through contact-less gestures, the performer can interact with physical instruments mimicking a maestro conducting a real orchestra. The installation consists of a set of 24 singing bowls placed radially into two rows as depicted by Figure 1. Each bowl produces a different tone, resulting from mechanical hammer's rebound which can be stopped using a muffler. The visitor can take place in the middle of the artwork and its body gestures are gathered using a Kinect camera device. To mimic the performance of a maestro, we devise a set of both pointing and semaphoric gestures such as swipe, tap or arm opening.

The *virtual hammer* gesture is used as a directional tap gesture from any hand to sing a given bowl. Such gesture observe the acceleration to control the hammer hitting force. While the hand remains pointing to the same location, the pointed bowl is hit repetitively. Adjacent bowls can be hit successively by swiping the hand to its location. Using each hand, the visitor can sing several bowls concurrently. If the pointing gesture is performed with both hands to a singing bowl, the muffler is triggered stopping the bowl tone. If the performer opens their arms all singing bowls are stopped at the same time. Finally, when both hands are getting closer to each other in front of the player, the last notes are replayed as maestro musical partition.

4.2 Tempo Scaduto

The second artwork installation is the Tempo Scaduto which was created by Vincent Ciciliato . It consists of an interactive movie where the visitor can control the movie storyline performing gestures in front of the screen as illustrated in Figure 3. The synopsis of the movie tries to revive both historical mafia facts and crime scenes. The visitor can access such information by playing the role of an assassin choosing their targets to evolve between historical facts while following the storyline proposed by the artist. The

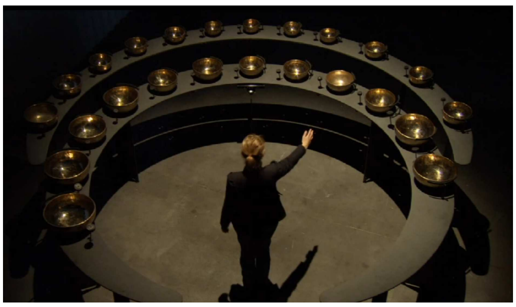

Figure 2: The artist mimicking a maestro using its own artwork installation.

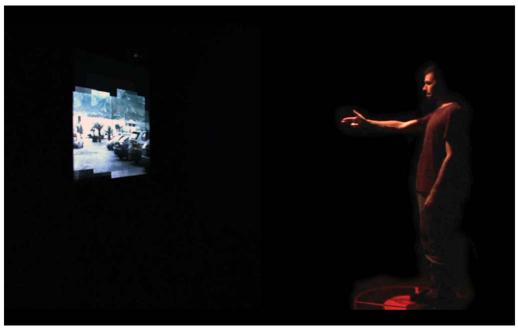

Figure 4: The Tempo Scaduto author demonstrating his work during an exhibit.

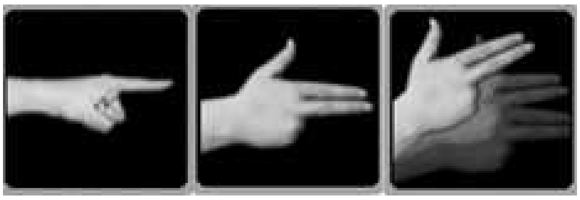

Figure 5: The three hand states used by the Tempo Scaduto installation.

revival of the scenes is presented using panoramic like views where people appear and disappear. By mimicking a gun shot with their hands, the visitor accesses movie sequences explaining historical facts. Due to the characteristic of the gesture, two cameras in the installation allow the tracking of both visitor body and their hand gestures. First, a Kinect camera is placed in front of the visitor to track their arms direction. A second camera is placed aside the user looking to their hand shape to recognize specific poses as shown in Figure 5. The *virtual gun* gesture combines observers using both cameras and takes advantage of the affordance provided by the handgun shape pose related to the historical context of this interactive artwork as illustrated in Figure 4.

4.3 Discussion

Both interactive art installations were designed favoring the usage of pantomimic gestures familiar to the visitors in the thematic of the artwork, i.e. mimicking a maestro or a handgun child game.While we placed a notice to describe supported gestures on both installations, most of the participants preferred to observe previous visitors before taking part of the proposed interactive experience. Visitors seem to enjoy the installation taking pictures of them self while gesturing. The gestural interaction, without requiring to instrument the visitor to track its body, provides a new dimension to the artwork and to the user experience even if the thematic of the second installation might dissuade them.

Regarding our gestural based approach, it demonstrated to be flexible enough to propose a wide range of gestures from a basic set of observers. The modular approach provided by the observer pattern and its composition allows to easily define new gestures as well as it easy the integration of different approaches . For example, the second installation relies on two distinct recognition techniques and it was done by two separate research team: one specialized in vision based algorithms and the other on 3D tracked skeleton analysis. In addition, it allows to handle the interactive requirement of each installation.

Finally we noticed problems inherent to gesture recognition. First, the visitors were getting closer and closer to the camera when the recognition was not behaving correctly, making their body tracking even harder to achieve. The different granularity requirements of both hand pose detection and body tracking can be an issue as seen by the setup of the second artwork installation.

5. CONCLUSION AND FUTURE WORK

In this paper, we introduced a tool adapted to contactless interaction to represent parametrized gestures and a new contactless interaction technique for clicking. We validated all of them in art installation from which we collected interesting observations. As future work, a study for confirming or improving our parameter classification is a direct continuation of our reflection to adapt existing taxonomies to parameterized gesture.Our toolkit for easing interaction design implements basics for contactless and full-body interaction. However a graphic interface, similar to ICON and Flowstate, is an interesting next step.

6. REFERENCES

[1] R. Aigner, D. Wigdor, H. Benko, M. Haller, D. Lindbauer, A. Ion, S. Zhao, and J. T. Koh. Understanding mid-air hand gestures: A study of human preferences in usage of gesture types for hci. *MSR-TR-2012-111*, 2012.

[2] C. Appert, S. Huot, P. Dragicevic, and M. Beaudouin-Lafon. FlowStates: prototypage d'applications interactives avec des flots de données et des machines à états. In *Proc. of AF-IHM*, pages 119–128, France, 2009.

[3] L.-W. Chan, H.-S. Kao, M. Y. Chen, M.-S. Lee, J. Hsu, and Y.-P. Hung. Touching the void: Direct-touch interaction for intangible displays. In *Proc. of CHI '10*, pages 2625–2634. ACM, 2010.

[4] A. Choumane, G. Casiez, and L. Grisoni. Buttonless clicking: Intuitive select and pick-release through gesture analysis. In *Proc. of IEEE VR 2010*, pages 67–70, 2010.

[5] P. Dragicevic and J.-D. Fekete. Input device selection and interaction configuration with icon, 2001.

[6] S. Gustafson, D. Bierwirth, and P. Baudisch. Imaginary interfaces: Spatial interaction with empty hands and without visual feedback. In *Proc. of UIST '10*, pages 3–12, NY, USA, 2010. ACM.

[7] R. Jota, M. A. Nacenta, J. A. Jorge, S. Carpendale, and S. Greenberg. A comparison of ray pointing techniques for very large displays. In *Proc. of GI 2010*, pages 269–276. Canadian Information Processing Society, 2010.

[8] M. Karam and M. C. Schraefel. A taxonomy of gestures in human computer interactions. Technical report, 2005.

[9] J. S. Pierce, A. S. Forsberg, M. J. Conway, S. Hong, R. C. Zeleznik, and M. R. Mine. Image plane interaction techniques in 3d immersive environments. In *Proc. of I3D 1997*, page 39. ACM, 1997.

[10] H. Rateau, L. Grisoni, and B. De Araujo. Mimetic interaction spaces : Controlling distant displays in pervasive environments. In *Proc. of IUI'14*, 2014.

[11] Y. Rekik, L. Grisoni, and N. Roussel. Towards many gestures to one command: A user study for tabletops. In *Proc. of INTERACT 2013*, volume 8118 of *LNCS*, pages 246–263. Springer, 2013.

[12] G. Ren and E. O'Neill. 3d selection with freehand gesture. *Computers & Graphics*, 37(3):101 – 120, 2013.

[13] M. Schwaller, S. Brunner, and D. Lalanne. Two handed mid-air gestural hci: Point + command. In *Proc. of HCI Interaction Modalities and Techniques*, volume 8007, pages 388–397. Springer, 2013.

[14] D. Vogel and R. Balakrishnan. Distant freehand pointing and clicking on very large, high resolution displays. In *Proc. of UIST'05*, pages 33–42. ACM, 2005.

[15] A. D. Wilson and A. F. Bobick. Nonlinear phmms for the interpretation of parameterized gesture. In *Proc. of IEEE Computer Vision and Pattern Recognition 1998*, pages 879–884. IEEE, 1998.

[16] A. D. Wilson and A. F. Bobick. Recognition and interpretation of parametric gesture. In *Proc. of IEEE Computer Vision 1998*, pages 329–336. IEEE, 1998.

[17] J. O. Wobbrock, M. R. Morris, and A. D. Wilson. User-defined gestures for surface computing. In *Proc. of CHI'09*, pages 1083–1092. ACM.